GRID-SEARCH LOCATION METHODS FOR GROUND-TRUTH COLLECTION FROM LOCAL AND REGIONAL SEISMIC NETWORKS

William Rodi,[1] Craig A. Schultz,[2] William G. Hanley,[2]
Sudipta Sarkar,[1] and H. Sadi Kuleli[1]

Massachusetts Institute of Technology,[1] Lawrence Livermore National Laboratory[2]

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. DE-FC03-01SF22397[1] and W-7405-ENG-48[2]

ABSTRACT

This paper reports on preliminary efforts to develop improved seismic event location techniques that can be used to generate more and better quality reference events using data from local and regional seismic networks. Our approach builds on earlier and concurrent work on the development of grid-search-based algorithms for single-event location and multiple-event location (known as GSEL and GMEL, respectively). Owing to the flexibility of the grid-search approach, we are now able to address certain additional complexities that we expect to yield improved reference event locations. These are (1) allowing spatially variable (source-dependent) station travel-time corrections in a multiple-event location analysis, and (2) extending the statistical model of observational errors beyond the simple class of Gaussian and non-Gaussian models now used.

On the first topic, we have formulated an extension of the multiple-event location algorithm GMEL to be a joint location and kriging method. The purpose of this is to allow the simultaneous processing of non-clustered, spatially well-distributed events. The kriging paradigm is extended such that a universal function, or set of functions, generates travel-time corrections at all stations. This allows the incorporation of important physical constraints such as source-receiver reciprocity. On the second topic, we are developing a new probabilistic error model in terms of mixture-of-Gaussians (MOG) distributions. Such distributions can describe multiple error processes that are not adequately described with unimodal, symmetric distributions.

OBJECTIVE

The objective of this work is to develop improved earthquake location techniques that can be used to increase the number and accuracy of reference events obtained from local and regional seismic networks. Our approach is to extend existing multiple-event location methods to incorporate improved statistical models of the observational errors in the arrival times measured from low-magnitude events, and spatially variable path corrections that allow the processing of non-clustered events.

RESEARCH ACCOMPLISHED

The approach of multiple-event location has been successfully applied as a tool for expanding the set of well-located earthquakes that can be used as reference events for the calibration of seismic monitoring stations (Engdahl and Bergman, 2000; 2001). The key feature of this approach is that it pools information from several events recorded at various subsets of a seismic network, together with ground-truth (GT) constraints on one or more of the events (e.g. from a dense local network), with the result that the location accuracy of the events is improved over that obtained by locating them independently. When the location error is reduced sufficiently (e.g. <5 km), an event becomes a useful reference event for calibration.

In conjunction with another project, we have extended an earlier grid-search event location algorithm for locating single events (Rodi and Toksöz, 2000) to a basic multiple-event location algorithm applicable to small event clusters. The algorithm performs the same basic task as the hypocentral decomposition (HDC) method of Engdahl and Bergman (2000, 2001) and as several other multiple-event location algorithms (see Rodi *et al*., 2002), except that it accommodates a certain class of non-Gaussian models of observational errors. Our algorithm for multiple-event location, known as GMEL, is described in Rodi *et al*. (2002). Here, we summarize what GMEL does and our progress in designing the extensions to this algorithm needed for this project.

Grid-Search Multiple Event Location

GMEL simultaneously processes seismic arrival time, azimuth and slowness data observed for multiple events, stations and phases. For simplicity, we consider here the special case in which the data set comprises only arrival times and assume that at most one arrival (P) per event has been observed at a station. Let d_{ij} denote the arrival time observation for the ith station and jth event, where $i = 1, \ldots, n$, and $j = 1, \ldots, m$. The multiple-event location problem addressed by this project can be written

$$d_{ij} = t_j + T_i(\mathbf{x}_j) + c_{ij} + e_{ij} \, , \tag{1}$$

where t_j and $\mathbf{x}_j$, respectively, are the origin time and hypocenter of the jth event; T_i is the travel-time function for the ith station, obtained from an assumed Earth model; c_{ij} is an unknown path correction; and e_{ij} is an observational ("picking") error. Equation (1) applies only to the i, j pairs for which arrival times have been observed.

GMEL currently assumes that the data errors are random and independent, and that each has a generalized Gaussian distribution of order p, whose probability density function (p.d.f.) is given by

$$f(e_{ij}) = \frac{1}{K(p)\sigma_{ij}} \exp\left\{ -\frac{1}{p} \left| \frac{e_{ij} - \mu_{ij}}{\sigma_{ij}} \right|^p \right\} \tag{2}$$

(see Billings *et al*., 1994). In this formula, μ_{ij} is the mean of the p.d.f., σ_{ij} is its standard error, and $K(p)$ is a constant depending on p. The order p can be any number greater than or equal to one. For $p = 1$, the p.d.f. is a Laplace distribution (two-sided exponential) and for $p = 2$, it is Gaussian. GMEL currently assumes that the μ_{ij} are known (usually zero) and that the σ_{ij} are known in a relative sense, whereby

$$\sigma_{ij} = \sigma_i \, \nu_{ij} \tag{3}$$

with the ν_{ij} being known but the station-dependent scale parameters, σ_i, being unknown.

Regarding the path corrections, GMEL currently assumes that they are event-independent, implying

$$c_{ij} = c_i \tag{4}$$

where the c_i are station-dependent travel-time corrections. This assumes, in effect, that the events are in a small cluster.

GMEL jointly solves for the problem unknowns: t_j, $\mathbf{x}_j$, c_i and σ_i. It solves a maximum-likelihood criterion, i.e. it maximizes the likelihood function determined by the assumed error p.d.f.'s. Its algorithm for maximizing likelihood combines grid-search (for the $\mathbf{x}_j$), root-finding (for t_j and c_i) and analytical (for σ_i) techniques. (Root-finding to obtain t_j and c_i is used when $p \neq 2$.) The likelihood function is maximized subject to specified upper and lower bounds on each parameter, which in the case of the event parameters allows the incorporation of GT information.

We point out that with the generalized Gaussian error model, maximizing likelihood with respect to the event locations and station corrections, with the σ_i fixed, is equivalent to minimizing the data misfit function given in terms of the L_p norm of the data residuals (see Rodi *et al*., 2002).

Examples of Multiple-Event Location

We applied GMEL and our single-event grid-search location algorithm (GSEL) to events from the Izmit/Duzce earthquake sequence reported in the International Data Centre (IDC) Reviewed Event Bulletin. The data set comprised 205 events in the area of the North Anatolian Fault (NAF) within the time period 17 August 1999 (Izmit main shock) through 26 February 2000. Only P, Pg and Pn arrivals were used, which numbered 2471 at 44 International Monitoring System (IMS) stations. First, we applied GSEL to locate each event independently. Then we applied GMEL to locate the events jointly with the determination of station corrections. The corrections were bounded between -10 and 10 seconds (essentially unconstrained) and the station-dependent standard errors (σ_i) were bounded between 0.5 and 1.5. GMEL was run with both a Gaussian error model and Laplace error model (p = 1). Event depths were bounded between 0 and 300 km.

The results are shown in Figures 1 and 2. The improvement between single and multiple event location is most evident for events recorded at more than a few stations. To show this, the first figure displays only the 114 events with 10 or more arrivals, and the second one displays the 60 events with 15 or more arrivals. Both figures show that, compared to GSEL (top panels), the locations determined with GMEL exhibit a much more linear pattern that roughly parallels the surface expression of the North Anatolian Fault. Further, the spatial pattern of epicenters is slightly better when a Laplace error model (bottom panel in each figure) is used instead of a Gaussian error model (center panel).

We point out that, in this application, the assumption of event-independent station corrections is probably not very accurate since the event cluster has an east-west aperture of about 250 km while the closest station is within this distance of most of the events. However, the value of the multiple-event approach is still apparent; below we will discuss our efforts to relax the small-cluster restriction.

Extended Error Model

Measurements of the onset time, azimuth and slowness of seismic arrivals are subject to a number of different error processes, and often result in heavy-tailed and asymmetric residual distributions. Numerous studies of arrival times, going back to Jeffreys (1932), have tried to understand and remedy the impact these complex error processes have on the event locations. Buland (1986) and Engdahl *et al*. (1998), for example, used heavy tailed Cauchy distributions to remap phases and improve phase picks in global catalogs. However, only limited efforts have been made to incorporate distributions other than Gaussian into event location procedures, e.g. Billings *et al*. (1994), Dreger *et al*. (1998) and our current efforts with GSEL/GMEL. These efforts have all adopted the generalized Gaussian distribution, defined in equation (2), in their location procedures. While the generalized Gaussian can capture some of the complexities in picking error processes by allowing increasingly heavy tails for smaller p, it is still a symmetric, unimodal distribution that fails to capture problems of identifying the onset of low signal-to-noise signals and misidentifying secondary phases as the first arrival. Unless such phenomena are properly accounted for

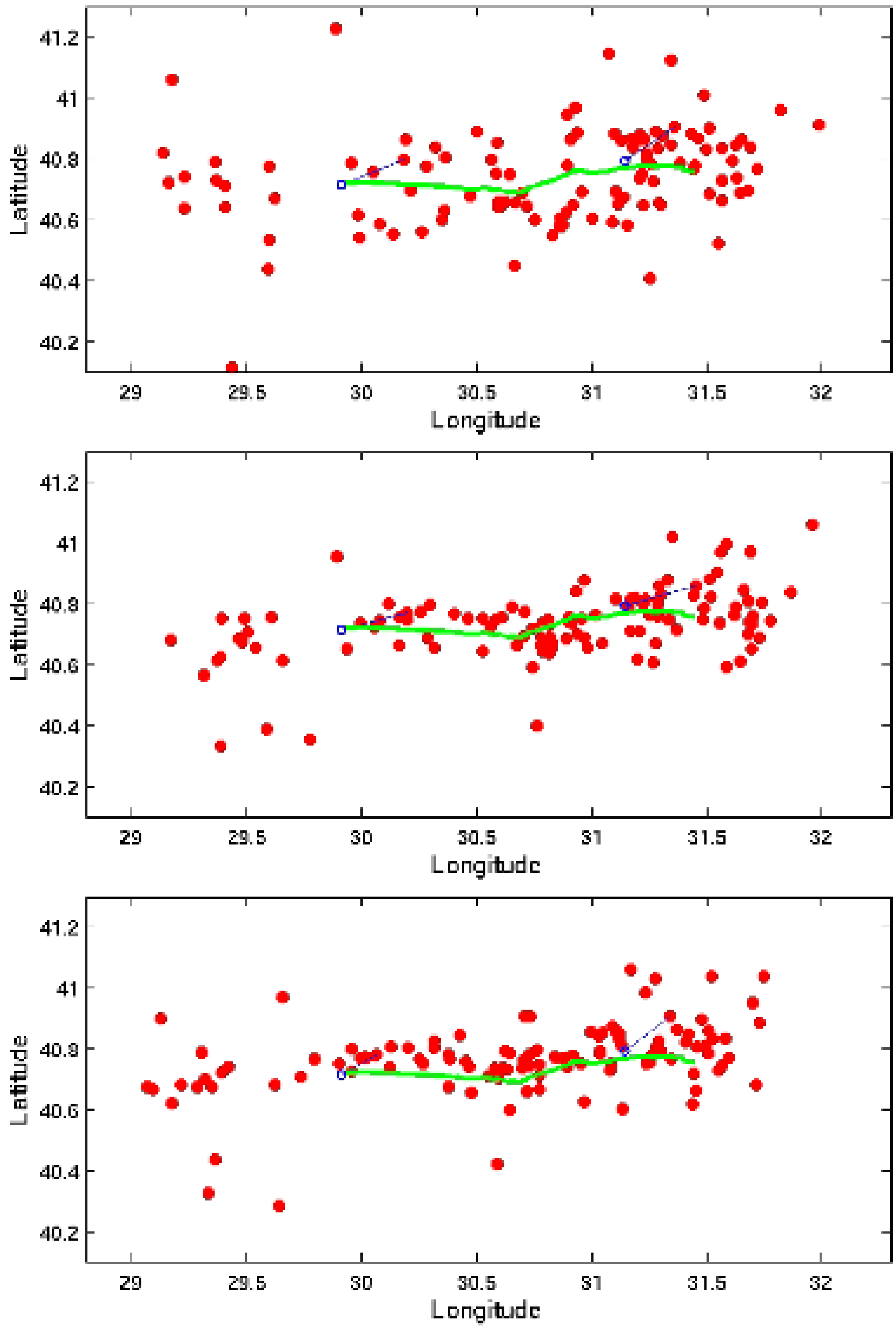

Figure 1: Locations of events from the Izmit/Duzce earthquake sequence, determined three ways. *Top*: Events located individually (GSEL). *Middle*: Multiple-event location (GMEL) with Gaussian (p = 2) error model. *Bottom*: Multiple-event location (GMEL) with Laplace (p = 1) error model. Only the events with 10 or more arrivals are shown. The continuous line is the surface expression of the North Anatolian Fault. The Izmit and Duzce mainshocks are connected to local network solutions (open circles) for these two events.

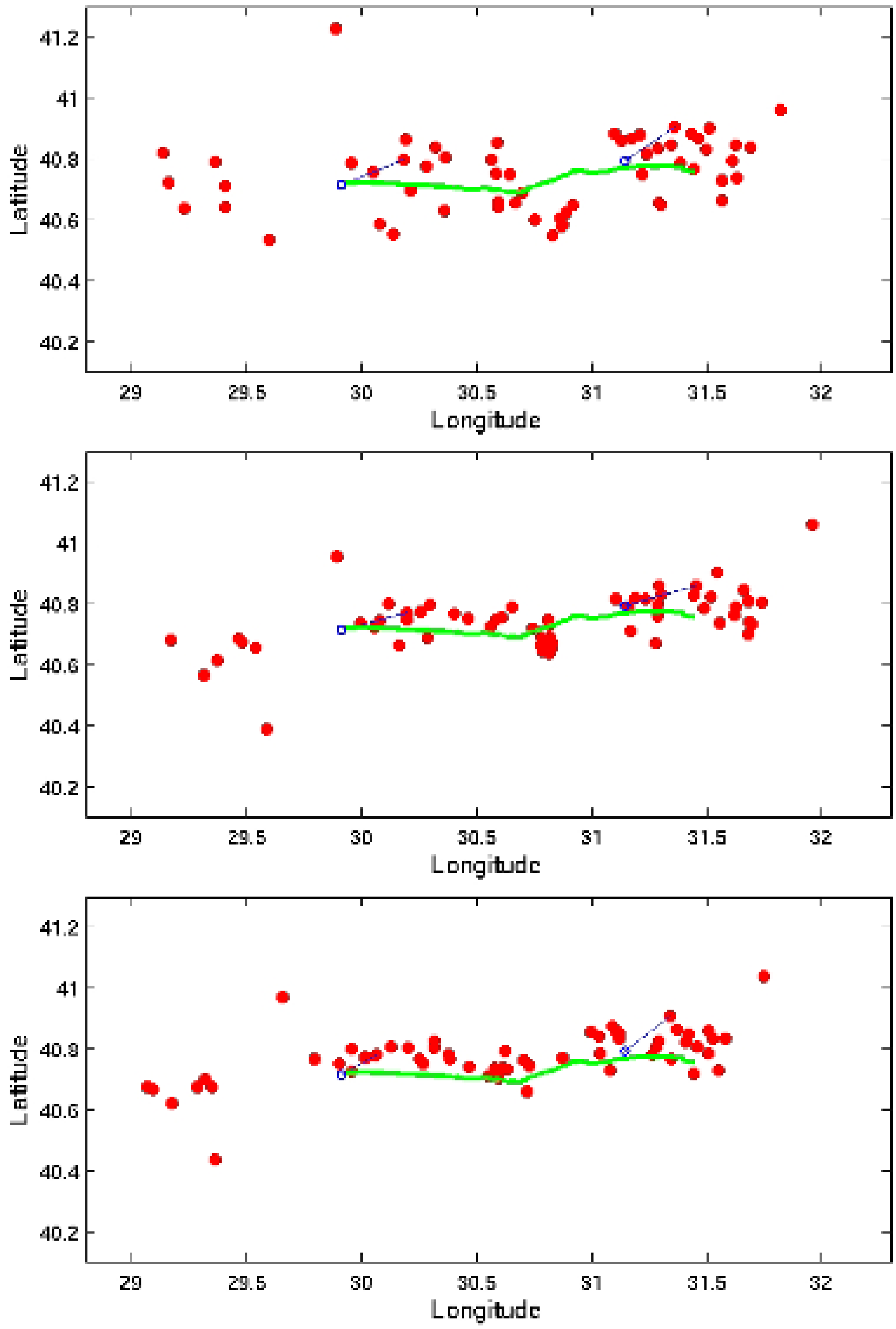

Figure 2: Locations of Izmit/Duzce earthquakes, determined three ways. This figure is the same as the preceding one except that only the events with 15 or more arrivals are shown.

in the location process, the errors in event locations obtained with sparse networks can be grossly underestimated. Figure 3 illustrates these difficulties with data from the 1991 Racha, Georgia, earthquake sequence.

In this project we are pursuing an alternative class of error distributions, first proposed by Jeffreys (1932), in which the error distribution is represented as a "mixture," or weighted sum, of Gaussian distributions having different means and variances. For K terms in the mixture, the p.d.f. of an observational error e is given by

$$f(e) = \sum_{k=1}^{K} a_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{ -\frac{1}{2}\left(\frac{e-\mu_k}{\sigma_k}\right)^2 \right\} \tag{5}$$

where the weights, a_k, sum to one. When the means, μ_k, differ, the distribution will be asymmetrical and allow, for example, a longer tail for positive errors (late picks) than for negative. Also, if the means differ sufficiently compared to the standard deviations, the distribution can be multi-modal.

We are currently pursuing the following tasks involved in implementing the mixture-of-Gaussians (MOG) error distribution in seismic event location. The first is the estimation of MOG distributions from observed travel-time residuals from regional and teleseismic stations. There are a variety of estimation methods that can be used for this purpose, but the most promising approaches we have identified are a variant of the expectation/maximization (EM) method (see McLachlan and Peel, 2000) and a Bayesian approach effected via a Gibbs or, more generally, Monte Carlo Markov Chain simulation (Gelman *et al*., 1995; Stephens, 1997). We are investigating these two approaches with respect to such issues as the need for good starting points, convergence rates and computational efficiency, divergence characterization (i.e., robustness of the process), and estimation error.

The second task we are pursuing is the modification of our grid-search multiple-event location algorithm (GMEL) to accept MOG error distributions. The use of MOG in place of generalized Gaussian p.d.f.s modifies the functional form of the likelihood function. While this poses no difficulty for a grid-search method, there are two issues we must deal with. One is that the root-finding procedure now used to maximize likelihood with respect to event origin times (t_j) and station corrections (c_i) must now deal with the fact that the likelihood may now be a multi-modal function of these parameters (not just the hypocenters). The second is that, while grid search already handles the multi-modality with respect to the $\mathbf{x}_j$, the possibility of local maxima will increase and the grid search may have to be that much more thorough.

Multiple-Event Location With Kriging

The assumption of source-independent station corrections restricts our multiple-event location method to events in clusters that are small compared to the distances to stations. This imposes a severe restriction on the event-station geometries that can be used in the multiple-event analysis. For example, stations too close to the cluster must be excluded and then processed separately to extract GT constraints, if an adequate local network exists for this purpose. A further restriction is that different clusters, even if they are reasonably near one another, must be processed separately, which makes it difficult to impose constraints on the spatial variability of path corrections as a function of source location.

To address these difficulties, we have formulated a generalization of the multiple-event location problem in which the path corrections are spatially variable. For example, if we replace the source-independent constraint in (4) with

$$c_{ij} = c_i(\mathbf{x}_j)\ , \tag{6}$$

the problem then becomes a joint inverse problem of finding the m event locations and n station correction *functions* (or "surfaces"), $c_i(\mathbf{x})$. The problem of estimating station correction functions has been addressed by Schultz *et al.* (1998) with the geostatistical interpolation method known as *kriging*. However, this procedure is currently applied with the event locations fixed, and then independently for each station.

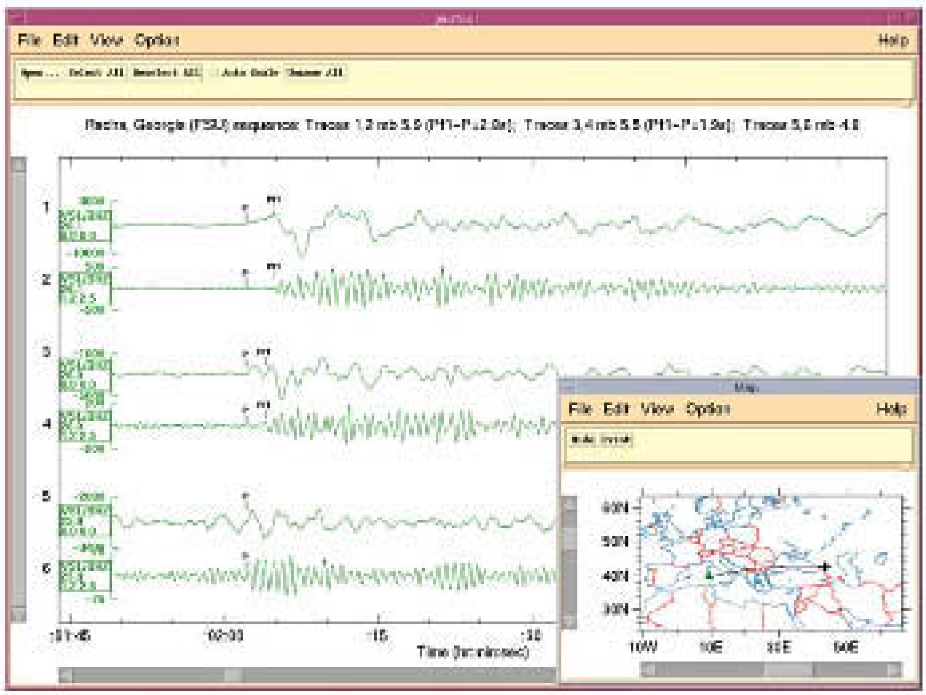

Figure 3. The Racha earthquake sequence exhibits the complex error processes in traveltime picks. This figure shows events of different magnitude filtered in different filter bands. The first two seismograms show the effect of magnitude on traveltime picks provided by a Livermore analyst. A large emergent first arrival causes about a 3 second shift from the pick on a regional event with less bandwidth. The second and third set of figures shows the misidentification of phases as the signal to noise decreases. The first smaller P phase (identified correctly for a large event) is missed on a regional event from the same source region. A secondary phase is misidentified and picked. This can lead to strong multi-modal behavior in the picking distribution and will be modeled with our technique.

Our formulation departs from the conventional kriging paradigm in the following ways. First, we define path corrections in terms of a universal correction function, or set of correction functions, which determine the path corrections for all events and stations. We also require the path corrections to obey source-receiver reciprocity. A simple example, which illustrates the concept but which may be too restrictive in practice, is to set

$$c_{ij} = \mathrm{c}(\mathbf{y}_i) + \mathrm{c}(\mathbf{x}_j) \qquad (7)$$

where $\mathbf{y}_i$ is the location of the ith station. The second departure from conventional kriging is to replace the minimum-variance criterion with a maximum-likelihood one. That is, we can jointly find event locations and, in the example of equation (7), a correction function $\mathrm{c}(\mathbf{x})$, by maximizing the likelihood function already defined in our maximum-likelihood formulation. However, since the likelihood depends on only samples of $\mathrm{c}(\mathbf{x})$, at station and event locations, it is necessary to add a term that regularizes the problem. If this term penalizes the spatial derivatives of $\mathrm{c}(\mathbf{x})$, it imposes the equivalent constraint as the correlation function used in conventional kriging. We are currently developing the numerical techniques to perform the likelihood maximization with respect to correction functions.

CONCLUSIONS AND RECOMMENDATIONS

While we are clearly in the algorithm development stage at this point of our project, we believe we have formulated appropriate extensions of the multiple-event location method that will allow the method to be applied to less restrictive source-receiver geometries and with more realistic models of the errors in seismic arrival time data. Therefore, our approach will potentially be able to obtain references events, suitable for calibration studies, of smaller magnitude and in new source areas. We are actively developing the techniques to implement these extensions.

REFERENCES

Billings, S.D., M.S. Sambridge and B.L.N. Kennett (1994), Errors in hypocenter location: picking, model and magnitude dependence, *Bull. Seism. Soc. Am.* **84**, 1978-1990.

Buland, R. (1986), Uniform reduction error analysis, *Bull. Seism. Soc. Am.* **76**, 217-230.

Dreger, D., R. Uhrhammer, M. Pasyanos, J. Franck and V. Romanowicz (1998), Regional and far-regional earthquake locations and source parameters using sparse broadband networks: a test on the Ridgecrest sequence, *Bull. Seism. Soc. Am.* **88**, 1353-1362.

Engdahl, E.R. and E.A. Bergman (2000), Identification and validation of reference events within the area being regionally monitored by IMS stations in Asia and North Africa, *Proceedings*, 22nd Annual DoD/DOE Seismic Research Symposium, New Orleans.

Engdahl, E.R. and E.A. Bergman (2001), Validation and generation of reference events by cluster analysis, *Proceedings*, 23rd Annual DoD/DOE Seismic Research Review, Jackson Hole, Wyoming.

Engdahl, E.R., R. van der Hilst and R. Buland (1998), Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seism. Soc. Am.* **88**, 722-743.

Gelman, A., J. Carlin, H. Stern and D. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, 420-438.

Jeffreys, H. (1932), An alternative to the rejection of observations, *Mon. Not. R. Astr. Soc, Geophys. Suppl*. **2**, 78-87.

McLachlan, G. and D. Peel (2000), *Finite Mixture Models*, Wiley Series in Probability and Statistics, 81--116.

Rodi, W. and M.N. Toksöz (2000), Grid-search techniques for seismic event location, *Proceedings*, 22nd Annual DoD/DOE Seismic Research Symposium, New Orleans.

Rodi, W., E.R. Engdahl, E.A. Bergman, F. Waldhauser, G.L. Pavlis, H. Israelsson and M.N. Toksöz (2002), A new grid-search multiple-event location algorithm and a comparison of methods, this volume.

Schultz, C.A., S.C. Myers, J. Hipp and C.J. Young (1998), Nonstationary Bayesian kriging: a predictive technique to generate spatial corrections for seismic detection, location, and identification, *Bull. Seism. Soc. Am.* **88**, 1275-1288.

Stephens, M. (1997), Bayesian methods for mixtures of normal distributions, Ph.D. Thesis, Mathematical Sciences, University of Oxford.